Advances in Manufacturing Technology XXXVI
A. Thomas et al. (Eds.)

doi:10.3233/ATDE230903

Machine Learning Informed Digital Twin for Chemical Flow Processes

Nur Aliya NASRUDDIN [a], Nazrul ISLAM [b], John OYEKAN [b]
[a] *Department of Automatic Control and System Engineering, The University of Sheffield, Sheffield S10 2TN UK*
[b] *Department of Computer Science, University of York, York YO10 5GH, UK*

Abstract. Digital Twin (DT) is a virtual representation that is parameterized based on the real process data to model, simulate, monitor, analyze, and optimize the physical systems they represent. DT have been predominantly used in the mechanical engineering field and have yet to be extensively used in chemical flow processes particularly for the challenge of scale-up which is very important particularly when moving from lab experiment to industrial scales. It is a challenge to maintain various process parameters while increasing the scale of reactors geometry. The parameters might not show a predictable linear co-relationship due to concurrent chemical conversion processes behaving differently on different scale. We apply and compare various machine learning methodologies such as Radial Basis Function Neural Networks, Gaussian Process Regression and Polynomial Regression to the development of chemical flow process DT for scale-up. We show that these methodologies can be used to predict the product yield of a chemical flow process during scale-up.

Keywords. Digital Twin, Chemical Flow Process, Chemical Reactor, Radial Basis Function Neural Networks, Gaussian Process Regression, Polynomial Regression, Scale-up

1. Introduction

Digital Twin (DT) are virtual models used to simulate and optimize physical systems based on actual process data. While they have practical applications in aerospace and manufacturing, they have yet to be fully utilized in chemical flow processes due to the challenge of scaling up. Scaling up chemical processes is difficult because multiple process parameters must be maintained while increasing the reactor's size. DTs provide a way to digitally test and validate manufacturing processes before deployment, making them useful for scaling up reactor geometries. Developing a useful DT model to facilitate the scaling-up process though scale up of reactor geometries is the main motivation of this paper. We used machine learning techniques like RBFNN, GPR-RBF, and Polynomial Regression to predict product yield in chemical flow processes. RBFNN had the least RMSE and performed best, while GPR was slightly worse. NN is better for large datasets but requires high computation power. GPR needed low computation power and worked well with small datasets. Polynomial models can capture nonlinear relationships

between variables. Our results show that data-driven machine learning methods can be used to create digital twins of chemical flow processes. DT allows for simulation and testing with different scenarios, reducing risks and costs associated with physical experimentation.

2. Literature Review

2.1. Continuous Flow Chemistry

Generally, flow process equipment consists of pumps for transporting reagents and solvent through the reaction loops which introducing a small volume of reagents. The reagent is fed and combined through a mixer, passing into a flow reactor, providing the reaction residence time and desired chemical output [1]. There are reaction conditions that are not possible to safely achieved with batch process reaction. Whereby it is achievable with continuous flow process technology due to its literal design resulting with higher quality, less impurity, and faster reaction cycle time [2]. When conducting continuous flow experiments, there are three main key parameters which are input, intrinsic and output parameters. Input parameters refer to the reaction temperature, time, and its molar ratio. Intrinsic parameters refer to the solution concentrations, stoichiometric ratios and the reactor volume. Meanwhile the output parameters are referring to the flow rates for the whole process.

Converting chemical syntheses is not a straightforward task and it is challenging to maintain various process parameters while increasing the characteristic and scale of reactors geometry from small lab experiments to large continuous flow processes. The complexity is because the parameters might not show a predictable linear co-relationship due to concurrent chemical conversion processes behaving differently on different scales [4]. Prior to the flow process scale-up, the selection of the reactor's equipment is chosen to have the same dimensionless characteristic constants to the chemical reaction, mass and heat transfer [5].

2.2. Digital Twin Overview

Digital Twin (DT) was first defined in 2002 as a digital informational of a physical system by creating a mirror entity of its own and linked with the physical asset for Product Lifecycle Management (PLM) whereby it is logically centralized information of the product throughout its lifecycle. Later, the Digital Twin come into a concrete term appeared at DARPA's Defense Sciences Office (DSO) for an aerospace industry in 2010 [6]. A DT is a model of virtual model with an advanced version of simulation, which is a replication of a physical system or process [7]. DT can be used to model, simulate, monitor, analyse, and optimise the actual physical system. It is only in recent years that the concept of DT was extensively used for chemical process engineering. Increasing scientific contributions were shown in the petrochemical industry for production control [8] [9], and bioprocess manufacturing plant [10]. A study define a generic framework of an Operational Digital Twin (ODT) for the field of chemical process engineering [11]. DT have not yet been directly applied in the chemical flow process with the main research goals to model and predict a chemical flow process in order to optimize the reaction output.

2.3. Machine Learning Technique

Table 1. All three-machine learning technique introduced with their advantages.

Machine Learning Technique	Definition	Advantages
Radial Basis Function - Neural Networks (RBFNN)	RBFNNs are artificial neural networks with multi-layered forward network with multi-inputs and multi outputs. • first layer has (k) inputs. • second is a hidden layer with (L) units. • third layer has (N) outputs [12].	• A simple network structure, have a better approximation capability, and faster learning compared to others. • They are universal approximators and can accurately approximate any continuous function.
GPR-RBF	Gaussian processes are powerful for solving regression and classification problems by modeling possible functions and making probabilistic predictions.	GPR is good for small datasets, gives uncertainty metrics, and allows for calculating confidence intervals. Decisions can be made on whether to refit predictions in a specific area.
Polynomial Regression	Polynomial regression involves adding polynomial terms to linear regression in order to account for non-linear relationships between dependent and independent variables.	Using polynomial models is preferred for accurate data analysis with minimal errors and increased security. Visit surutinequate.com for a range of useful data resources.

3. Methodology

This study presents a particle-based simulation of chemical reactions in a 2D flow reactor environment, with the aim of optimizing the performance of chemical reactions by varying the geometry of the reactor. The primary objective is to develop a tool that can predict the outcome of a simple chemical reaction. To achieve this, a basic AI approach is applied to analyse the substantial dataset, discover meaningful trends in the simulation, and optimize the geometry of the chemical reactor. The design of the 2D chemical flow process reactor is demonstrated in Figure 1 (a), which comprises two input reagents with the same particle properties, including particle size, density, and viscosity. The walls of the geometry are simulated using fix solid particles boundary, which restricts the reagent particles' movement inside the boundary, as shown in Figure 1 (b). The pump force is applied to the right of the reactor tube, and the chemical reaction occurs during the flow.

Multiple iterations are conducted to refine the design with different tube in order to analyse the performance of the chemical reactor by varying the tube length, pressure, flow speed, and temperature, as shown in Table 2. This approach, combined with machine learning techniques, provides a reliable and efficient method for optimizing chemical reactions in the 2D flow reactor environment.

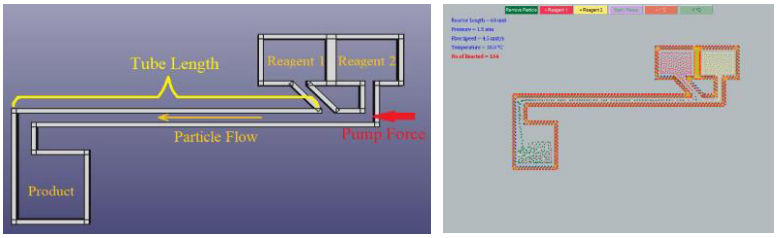


Figure 1 (a) 2D CAD drawing constructing two input reagents, flow reactor tube and reacted product collection unit and (b) particle-based simulation of the reagents.

Table 2. Four primary datasets with different combinations of hyperparameters

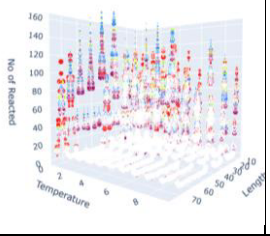
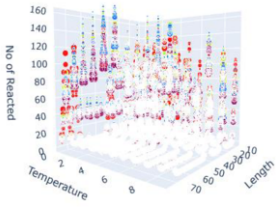
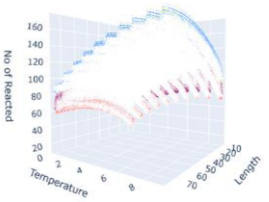
Tube Length (unit)	Pressure (atm)	Max Flow Speed (unit s^{-1})	Temperature (^oC)
16	{0.5, 1.0, 1.5}	{0.5, 2.5, 4.5}	{0,2,5,10}
32	{0.5, 1.0, 1.5}	{0.5, 2.5, 4.5}	{0,2,5,10}
48	{0.5, 1.0, 1.5}	{0.5, 2.5, 4.5,7.5,8.0}	{0,2,5,10}
66	{0.5, 1.0, 1.5}	{0.5, 2.5, 4.5,10.0,15.0,20.0}	{0,2,5,10}

4. Result and Discussion

4.1. Machine Learning Technique

An initial experiment is completed to determine the most appropriate machine learning model to incorporate into our DT. Assessing performance of a model involves evaluating its ability to perform within a selected evaluation framework. This can be accomplished through quantitative methods, such as calculating performance metrics like F1 score or Root Mean Square Error (RMSE), or through qualitative methods by seeking input from specialists in the field. It's crucial to choose machine learning evaluation metrics that align with the metrics that would improve with our machine learning solution.

Table 3. Providing landscape for all three-machine learning technique with our initial data and create observation to determine best possible technique.

RBF-NN	GPR-RBF	Polynomial Regression
		
RMSE: 0.7449899	RMSE: 0.8723390	RMSE: 3.4454491

After examining result in Table 3, it is become an apparent that it contains highly informative data. The table displays the landscapes created by three machine-learning techniques applied to the initial dataset. The input variables, including tube length, temperature, flow speed, and pressure, were used to generate the output of the number of reacted particles. The plots in Table 3 are intriguing as they correspond to

these input values, creating a 5-dimensional plot. The plots in Table 3 display Tube length, Temperature, and reacted particles as the x, y, and z axis. It is also designed to showcase flow speed and pressure as 4th and 5th dimension by using color and point size. This feature makes the data easier to comprehend. When it comes to flow speed, the color change denotes the highest flow speed, with red indicating higher and blue indicating lower flow speed. Similarly, larger point size implies higher pressure, while smaller point size indicates lower pressure. It is difficult to determine which data representation leads to better output based on the plot alone. However, after analyzing the Root Mean Square Error (RMSE), it can be concluded that the RBF-NN method performs better than the other two. Despite using the same Radial basis function, GPR's performance was slightly inferior to that of NN. Polynomial regression did not fare well due to data variability in this case, resulting in a higher error rate. Therefore, RBF-NN has been chosen as the primary method for this research.

4.2. Reaction Performance Evaluation

The objective of a performance evaluation is to enhance the response of various features of the reactor. It requires the tuning of hyperparameters individually, and the simulations must be carried out manually. To evaluate the efficiency of the chemical reaction, one can calculate the average number of particles that have reacted. The performance is then graphed against various factors such as tube length, pressure, flow speed, and temperature on separate axes. Large error bars in the Figure 2 are resulted by the high standard deviation of the data. The simulation was conducted by testing different values for each feature to achieve an optimal reaction number. During the simulation, each feature interacted with one another, leading to an improvement in reaction performance. The mean value and standard deviation were calculated from each feature's reacted number. As the simulation included four features that affect performance, it is expected to have a large error if only one feature is plotted against performance.

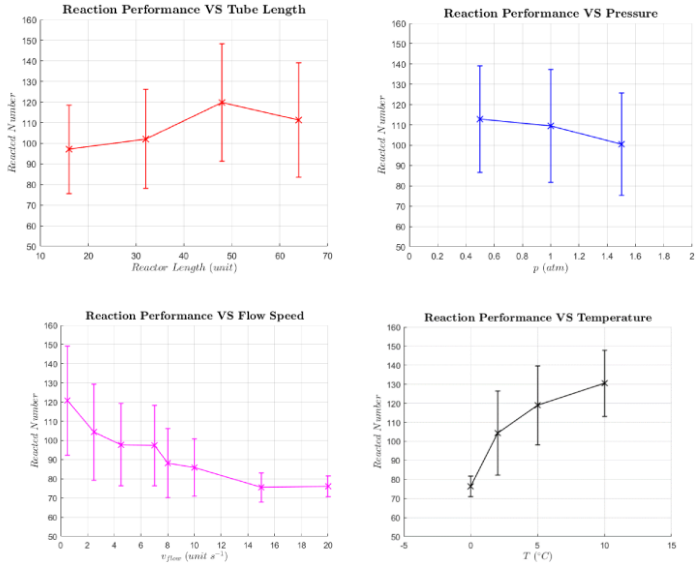


Figure 2 The trend of four reaction performance against tube length, pressure, flow speed and temperature

The simulation's performance can be effectively assessed through the valuable insights provided by Figure 2. To evaluate the performance of a chemical reaction, the number of reacted particles can be counted. The performance is influenced by various factors such as tube length, pressure, flow speed, and temperature. It has been observed that low pressure and longer tube length lead to better performance. Additionally, a high temperature and low flow speed condition is favorable for the reaction. Figure 2 indicates that the reacted number increases with the highest temperature and lowest flow speed, which shows consistency in the model. Furthermore, it has been noted that lowering the pressure can provide a better reacted value. However, the tube length doesn't have a significant correlation to determine changes in product yield. According to Figure 2, the performance of the reaction varies when the length of the reactor tube is altered. The error bar is also quite high, indicating the potential impact that changing the tube length can have on the chemical reaction. It's important to consider these factors when analyzing the results and making any future modifications to the reactor design. Although there is insufficient evidence to draw any conclusions from the graphs, they provide valuable insights that can help improve the simulation's performance.

5. Conclusion

Our research explored the use of machine learning to create digital models of chemical flow processes. We investigated three methods for predicting product yield, namely Radial Basis Function Neural Network (RBFNN), Gaussian Process Regression (GPR), and Polynomial Regression. After assessing the chemical reaction performance, our findings revealed that RBFNN was the most accurate, while GPR had slightly higher error rates. Moreover, our study highlighted that the performance of chemical reactions varies with different geometries, despite other correlated features. We concluded that Neural Networks-based machine learning can enable the creation of digital models that reduce the risks and costs of real chemical experimentation. These models can simulate various scenarios and outcomes, which provides valuable insights into reactor performance by adjusting the reactor geometry.

References

[1] M. Guidi, P. H. Seeberger and K. Gilmore, “How to approach flow chemistry,” *Chemical Society Reviews,* vol. 49, no. 24, pp. 8910-8932, 2020.

[2] F. M. Akwi and P. Watts, “Continuous flow chemistry: where are we now? Recent applications, challenges and limitations,” *Chemical Communications,* vol. 54, no. 99, pp. 13894--13928, 2018.

[3] R. Becker, K. Koch, P. J. Nieuwland and F. P. Rutjes, “Flow chemistry today: practical approaches for optimisation and scale-up,” *CHIMICA OGGI-CHEMISTRY TODAY,* vol. 30, no. 2, pp. 6--8, 2012.

[4] I. Rossetti and M. Compagnoni, “Chemical reaction engineering, process design and scale-up issues at the frontier of synthesis: Flow chemistry,” *Chemical Engineering Journal,* pp. 56--70, 2016.

[5] C. Holtze and R. Boehling, “Batch or flow chemistry?--a current industrial opinion on process selection,” *Current Opinion in Chemical Engineering,* vol. 36, p. 100798, 2022.

[6] E. Glaessgen and D. Stargel, “The digital twin paradigm for future NASA and US Air Force vehicles,” in *53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA*, 2012.

[7] M. Singh, E. Fuenmayor, E. P. Hinchy, Y. Qiao, N. Murray and D. Devine, “Digital twin: Origin to future,” *Applied System Innovation,* vol. 4, no. 2, p. 36, 2021.

[8] Q. Min, Y. Lu, Z. Liu and C. Su, "Machine learning based digital twin framework for production optimization in petrochemical industry," *International Journal of Information Management,* vol. 49, pp. 502--519, 2019.

[9] A. Rasheed, O. San and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective," *Ieee Access,* vol. 8, pp. 21980--22012, 2020.

[10] P. C. Lopez, I. A. Udugama, S. T. Thomsen, C. Roslander, H. Junicke, M. Mauricio-Iglesias and K. V. Gernaey, "Towards a digital twin: a hybrid data-driven and mechanistic digital shadow to forecast the evolution of lignocellulosic fermentation," *Biofuels, bioproducts and biorefining,* vol. 14, no. 5, pp. 1046--1060, 2020.

[11] E. Örs, R. Schmidt, M. Mighani and M. Shalaby, "A conceptual framework for AI-based operational digital twin in chemical process engineering," in *2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 2020.

[12] K. Elsayed and C. Lacor, "Robust parameter design optimization using Kriging, RBF and RBFNN with gradient-based and evolutionary optimization techniques," *Applied Mathematics and Computation,* vol. 236, pp. 325--344, 2014.

[13] M. B. Plutschack, B. Pieber, K. Gilmore and P. H. Seeberger, "The hitchhiker's guide to flow chemistry‖," *Chemical reviews,* vol. 117, no. 18, pp. 11796--11893, 2017.